

LAMP-TR-031
CAR-TR-905
CS-TR-3979

IRI-93-09100
IRI-97-12598
MDA9049-6C-1250
January 1999

Object Representation Using Appearance-Based Parts and Relations

Chien-Yuan Huang,¹ Octavia I. Camps,^{1,2} Tapas Kanungo³

¹Department of Electrical Engineering

²Department of Computer Science and Engineering
The Pennsylvania State University, University Park, PA, 16802
camps@whale.ece.psu.edu

³ Center for Automation Research
University of Maryland, College Park, MD 20742-3275
kanungo@cfar.umd.edu

Abstract

The recognition of general three-dimensional objects in cluttered scenes remains a challenging problem. In particular, the design of a good representation that is suitable for modeling large numbers of generic objects, and is also robust to occlusion, has been a stumbling block to achieving success. In this paper, we propose a representation using *appearance-based parts* and *relations* to overcome these problems. Appearance-based parts and relations are defined in terms of closed regions and unions of these regions, respectively. The regions are segmented using the MDL principle; their appearance is obtained from collections of images and compactly represented by parametric manifolds in the two eigenspaces spanned by the parts and the relations. Qualitative and quantitative experiments illustrating the potential of the representation for successful object recognition in the presence of clutter and occlusion are presented.

C. Y. Huang and O. I. Camps were supported in part by NSF grants IRI-93-09100 and IRI-97-12598. T. Kanungo was supported in part by the Department of Defense and the Army Research Laboratory under Contract MDA 9049-6C-1250.

| Report Documentation Page | | | | Form Approved OMB No. 0704-0188 | |
|--|------------------------------------|-------------------------------------|----------------------------|---|---------------------------------|
| Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. | | | | | |
| 1. REPORT DATE JAN 1999 | | 2. REPORT TYPE | | 3. DATES COVERED 00-01-1999 to 00-01-1999 | |
| 4. TITLE AND SUBTITLE Object Representation Using Appearance-Based Parts and Relations | | | | 5a. CONTRACT NUMBER | |
| | | | | 5b. GRANT NUMBER | |
| | | | | 5c. PROGRAM ELEMENT NUMBER | |
| 6. AUTHOR(S) | | | | 5d. PROJECT NUMBER | |
| | | | | 5e. TASK NUMBER | |
| | | | | 5f. WORK UNIT NUMBER | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Language and Media Processing Laboratory, Institute for Advanced Computer Studies, University of Maryland, College Park, MD, 20742-3275 | | | | 8. PERFORMING ORGANIZATION REPORT NUMBER | |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | | | | 10. SPONSOR/MONITOR'S ACRONYM(S) | |
| | | | | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) | |
| 12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited | | | | | |
| 13. SUPPLEMENTARY NOTES | | | | | |
| 14. ABSTRACT | | | | | |
| 15. SUBJECT TERMS | | | | | |
| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES 29 | 19a. NAME OF RESPONSIBLE PERSON |
| a. REPORT unclassified | b. ABSTRACT unclassified | c. THIS PAGE unclassified | | | |

LAMP-TR-031
CAR-TR-905
CS-TR-3979

IRI-93-09100
IRI-97-12598
MDA9049-6C-1250
January 1999

Object Representation Using Appearance-Based Parts and Relations

C. Y. Huang, O. I. Camps and T. Kanungo

Object Representation Using Appearance-Based Parts and Relations

Chien-Yuan Huang,¹ Octavia I. Camps,^{1,2} Tapas Kanungo³

¹Department of Electrical Engineering

²Department of Computer Science and Engineering
The Pennsylvania State University
University Park, PA, 16802
camps@whale.ece.psu.edu

³ Center for Automation Research
University of Maryland
College Park, MD 20742-3275
kanungo@cfar.umd.edu

Abstract

The recognition of general three-dimensional objects in cluttered scenes remains a challenging problem. In particular, the design of a good representation that is suitable for modeling large numbers of generic objects, and is also robust to occlusion, has been a stumbling block to achieving success. In this paper, we propose a representation using *appearance-based parts* and *relations* to overcome these problems. Appearance-based parts and relations are defined in terms of closed regions and unions of these regions, respectively. The regions are segmented using the MDL principle; their appearance is obtained from collections of images and compactly represented by parametric manifolds in the two eigenspaces spanned by the parts and the relations. Qualitative and quantitative experiments illustrating the potential of the representation for successful object recognition in the presence of clutter and occlusion are presented.

C. Y. Huang and O. I. Camps were supported in part by NSF grants IRI-93-09100 and IRI-97-12598. T. Kanungo was supported in part by the Department of Defense and the Army Research Laboratory under Contract MDA 9049-6C-1250.

1 Introduction

The recognition of general three-dimensional objects in cluttered scenes from two-dimensional images remains a challenging problem. In particular, the design of a good representation that is suitable for modeling large numbers of generic objects, and is also robust to occlusion, has been a stumbling block to achieving success.

A major difficulty in recognizing 3D objects from 2D images is that their appearances change significantly with the viewpoint. Common approaches to overcoming this problem are to use viewer-centered representations to describe the objects in terms of their appearances, or to use object-centered representations and image invariants.

Viewer-centered approaches can be as structured as features grouped into relational models within aspect views [6, 11], or as loose as appearance-based representations [29] constructed from collections of images. A major limitation of the appearance-based approach is that it requires isolating the object of interest from the background, and thus is sensitive to occlusion. In spite of the increased interest in this approach [24, 26], no satisfactory solution has been found, until now, to handle object occlusion without limiting the scale of the problem (the number of objects).

Approaches using object-centered representations, such as part decomposition [3, 42, 23], have the potential to cope with both occlusion and large object databases. However, the definition of parts of generic objects and their image extraction remains a difficult problem [17].

Dickinson *et al.* [13] proposed a hybrid approach where objects are described as combinations of geometric primitives that are represented using aspect graphs. This approach handles occlusion and can potentially describe a large set of objects in terms of a few primitives. However, it requires fairly good image segmentation and it is limited to objects that can be described by primitives of specific types.

In this paper¹, we propose a new method of describing the *appearance* of a 3D object using *parts* and *spatial relations* among these parts. This representation is automatically learned from a set of training images and is compactly stored as two sets of parametric manifolds, called Appearance-Based Parts (ABPs) and Appearance-Based Relations (ABRs). The ABP and ABR manifolds are embedded in low-dimensional subspaces of two eigenspaces spanned by collections of closed regions and unions of regions, respectively, segmented from the training data using the MDL principle. Since this representation is learned from segmented images, it is capable of representing free-form objects and handling segmentation problems similar to the ones encountered during training. Furthermore, since it is based on local regions rather than global properties, it is robust to partial occlusion.

The remainder of the paper is organized as follows. The next section discusses previously proposed object representations and gives the details of the new ABP and ABR representation. Section 3 presents a relational formalism to perform object recognition and pose estimation using the new representation. Section 4 describes experiments that illustrate, qualitatively as well as quantitatively, the usefulness of the new representation.

¹A shorter version of this paper appeared in [18].

2 Object Representation

The representation of models is critical to the problem of 3D object recognition and pose estimation. A good review of work in this area up to 1993 can be found in [20]. More recently, two workshops addressing the specific issue of object representation for recognition [17, 33] have shown that the problem of designing a “good” representation remains largely unsolved.

Most 3D object representations for recognition purposes proposed until now can be classified into three major categories: primitive-, physics-, and appearance-based representations.

Primitive-based representations rely on geometric models of objects [2, 12, 14, 19]. In order to cope with occlusion, decomposition of the object into parts is often used. Binford [3], for example, proposes a 3D part definition based on function. In this approach, a *divide and conquer* method is utilized to decompose complex objects into a structural representation. Another example can be found in [42], where Zerroug and Medioni employ a high-level, volumetric part-based approach where a hierarchical extraction process applies generalized cylinders to group compound objects from boundaries, surface patches, and volumetric parts. Although there is general consensus on the fact that part decomposition can help in overcoming occlusion, there is no agreement on what a part should be. Furthermore, reliable extraction of parts from 2D image data remains a difficult problem.

Physics-based representations typically model a shape as a mechanical system subject to forces reflecting material properties as well as smoothness and image constraints. These methods have been successfully used for modeling complex objects whose shapes may vary over time. Examples of this approach are the work of Metaxas [27] and Pentland and Sclaroff [32, 38]. Metaxas proposed a deformable model by integrating mathematical methods from geometry, physics and mechanics. In particular, he used Lagrangian mechanics to convert the geometric parameters of the solid primitive, the deformation parameters, and the six degrees of freedom of rigid-body motion, into generalized coordinates or dynamic degrees of freedom. Pentland and Sclaroff, on the other hand, used a finite element method, where the eigenvectors of the finite element model of the shape were employed to formulate the physical model. However, these methods are better suited for 3D object recognition from 3D data or 2D object recognition from 2D data.

The appearance of a 3D object in a 2D image depends on its shape, its reflectance properties, its pose in the scene, and the sensor and illumination characteristics. Earlier object recognition systems such as PREMIO [8, 9] and the system developed by Chen and Mulgaonkar [10] used synthetic image segmentations to learn probability models that were then used to recognize 3D objects using a Bayesian framework. The main limitations of these systems are that they require CAD models of the objects and that the simulations are not as realistic as they should be. Costa and Shapiro [11] and Pope and Lowe [34] have addressed these problems in part by learning segmentations using real images.

More recently, Murase and Nayar [29] proposed a parametric eigenspace representation for the learning, recognition, and pose estimation of rigid objects. In this approach large sets of images obtained by varying pose and illumination in small increments are

compressed using a Karhunen-Loeve expansion. Variations on object appearance due to translation and scaling, on the other hand, are taken care of by normalizing the image size using the bounding box of the object, at both training and recognition times. In [28] Mundy *et al.* presented an experimental comparison between the appearance-based method of Murase and Nayar (SLAM) and two geometric model-based recognition methods described in [36] (Lewis) and [44] (Morse). This study concluded that the two approaches complement each other. Appearance models have the advantage that they do not require formal models to describe objects, while geometric approaches rely on formal models to derive pose-invariant properties. The major drawbacks of using object appearance are that i) it is very sensitive to segmentation, in particular occlusion; ii) it does not lend itself well to object categorization; and iii) incidental variations in appearance such as texture or surface albedo must be modeled as separate objects. The major drawback of the geometric approach is that it is not robust to minor variations of the hard constraints imposed on the image geometry.

Lately, the appearance-based approach has received increasing attention [26, 24]. In particular, there has been a significant effort devoted to overcoming the problems caused by occlusion and background clutter. In spite of these efforts, no satisfactory solution has been found to handle occlusion without limiting the scale of the problem (number of objects). In [26], for example, a robust method of computing the coefficients to project an image into the parametric eigenspace is presented. This method extracts the coefficients by considering subsets of image points with a hypothesize-and-test paradigm and selecting the best hypothesis by using the MDL principle. As a result, the coefficients are robust to image outliers and in particular to occlusion. However, a major problem with this technique is that it cannot handle object translation and scaling. This is because this approach works only if the dimensions of the training and testing images are equal, and the pixel locations of the object do not change at recognition time. Unfortunately, occlusion has a direct impact on the object’s bounding box, preventing the use of image size normalization in this case. In [4] an extension of this technique using multiresolution matching was proposed to handle object scaling. Although this technique is successful in finding objects at different scales, it is very time-consuming. Krumm [24] proposed handling occlusion by using small neighborhoods as features. Although this technique can handle object translation, it also suffers from the scaling problem — i.e., it assumes that the object size in the image is the same at recognition and training time.

It should be noted that the representation proposed in this paper is related to the appearance-based representation proposed by Murase and Nayar. However, the use of appearance-based *parts* and *relations* improves the representation’s robustness to segmentation problems and occlusion without compromising scaling or the ability to handle free-form objects. This is accomplished by using local rather than global appearances. Furthermore, the proposed representation is well suited to grouping together “similar” parts into categories, allowing several objects to share parts, as shown in [7].

2.1 Parts from Images

It is commonly accepted that complex objects can be decomposed into simple parts. However, there is not much agreement on how to define what a *part* is. Several definitions

have been proposed in the past, including operational definitions (parts are what a part detector finds), view-based definitions (parts are defined by local image properties), and geometric definitions (parts are defined by 3D events) [17].

We believe that a definition of a part must take into account the segmentation algorithms that will be used to extract the parts from the images. In particular, we believe that a part definition should be used in the same way at the learning *and* recognition stages. Thus, we have opted for the following definition:

Parts are polynomial surfaces approximating *closed, non-overlapping* image regions that *optimally partition* the image in a *minimum description length* (MDL) sense.

The MDL principle is a formalization of Ockham’s razor: “the simplest model explaining the observations is the best.” We have chosen an MDL-based definition for the following reasons:

1. The MDL principle has a strong theoretical grounding;
2. Using MDL does not require arbitrary parameters, and thus parts can be extracted in a consistent manner;
3. The MDL objective function is formulated such that statistics are tested inside the regions and such that the resulting regions have homogeneous intensity (color or texture) properties;
4. Finally, algorithms using fast incremental computations are available [21].

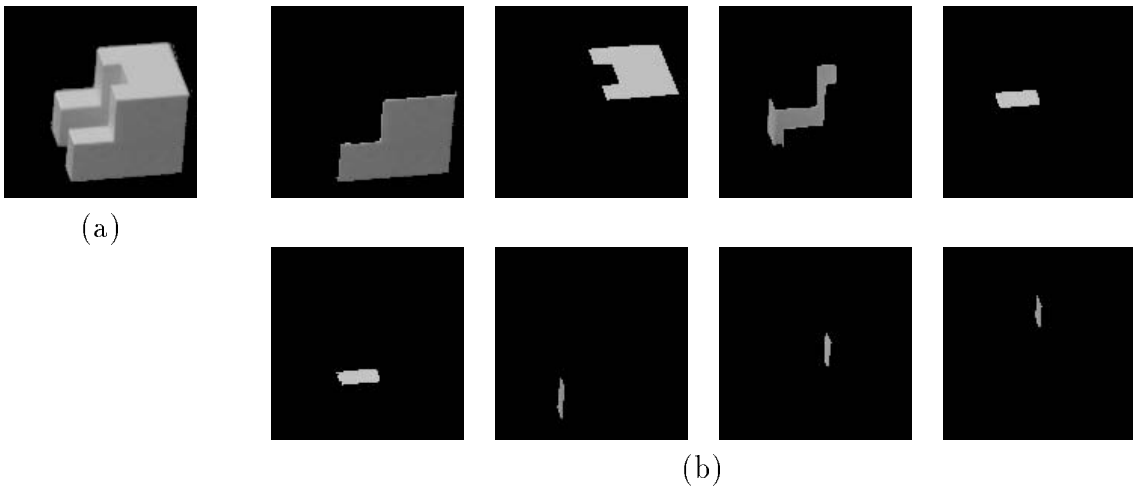


Figure 1: (a) Object "C-cube". (b) Parts obtained using an MDL-based segmentation algorithm.

The MDL principle has been applied to image segmentation [21, 25, 43] by setting up an *objective function* whose global minimum corresponds to the simplest description of

an image segmentation, and hence the best segmentation. The MDL objective function that we use is the one proposed in [21] resulting from describing an image segmentation as a collection of regions modeled as polynomial surfaces of variable degree, perturbed by zero-mean Gaussian noise and whose boundaries are encoded using a chain code representation. Although this model works best for constant-albedo regions, it can be easily adapted to textured regions by using a set of texture filters like the ones used by Zhu and Yuille in [43].

Let $\Omega = \{\omega_j\}$ denote the image segmentation into regions $\{\omega_j\}$ and let Y represent the image data. Further, assume that the image comes from a stochastic process that can be characterized as a polynomial gray scale surface of unknown degree plus Gaussian noise described by a vector of parameters β . Then, the MDL objective function to optimize is given by

$$L(Y, \Omega, \beta) = L(\Omega) + L(\beta|\Omega) + L(Y|\Omega, \beta). \quad (1)$$

where the first term is the length of encoding the region boundaries, the second term is the length of encoding the parameters, and the last term is the length of encoding the residuals. If the boundaries are encoded using their chain code representations, assuming that at each point the number of possible directions is 3 (i.e. the number of adjacent grid points, excluding the current one), the first term of the encoding cost can be approximated by [35]

$$L(\Omega) = \sum_i (l_i \log 3 + \log^*(l_i) + \log(2.865064))$$

where l_i is the length of the boundary i and $\log^*(x) = \log x + \log \log x + \log \log \log x + \dots$ (where the sequence stops when the terms become negative). The second cost term, $L(\beta|\Omega)$, can be expressed using Rissanen's [35] expression for optimal-precision analysis that says that K independent real-valued parameters characterizing n data points can be encoded using $(K/2) \log n$ bits. Thus,

$$L(\beta|\Omega) = \frac{1}{2} \sum_j K_{\beta_j} \log n_j$$

where K_{β_j} is the number of free parameters describing region j and is a function of the polynomial degree to be determined, and n_j is the number of pixels in region j . Finally, the third cost term $L(Y|\Omega, \beta)$ can be written using Shannon's theorem [1] as

$$L(Y|\Omega, \beta) = -\log p(Y|\Omega, \beta) = \sum_j -\log p(Y_j|\beta_j)$$

Figure 1(a) shows an image where the object "C-Cube" has been thresholded from the background, and Figure 1(b) shows the largest parts obtained using the MDL-based segmentation algorithm described in [21]. Each of the eight parts is shown in a separate image.

2.2 Appearances of Parts

Obviously, parts obtained using the definition given above are sensor- and illumination-dependent. Thus, in order to completely characterize an object for different sensors and light sources, we introduce the concept of "appearances" of a part:

Two parts segmented from two images of the same object obtained with similar sensor and illumination configurations, are said to be **appearances** of the same part if they are judged to have similar polynomial approximations in similar image locations.

This concept can be formalized as follows. Let ω_i be a part obtained from an image. Let Y_i be an $n_i \times 1$ column vector with the gray scale pixel values in part ω_i . Let d be the order of the polynomial used to fit the parts, and $m = (d + 1)(d + 2)/2$ be the number of polynomial coefficients. Let Φ_i be an $n_i \times m$ matrix of m basis functions for each of the n_i pixels — i.e., products of powers of pixel coordinates. Finally, let Θ_i be an $m \times 1$ column vector with the optimal regression coefficients for ω_i . Using these definitions, we have [21]

$$Y_i = \Phi_i \Theta_i + \Psi_i$$

where Ψ_i is a vector of zero-mean Gaussian noise with covariance $\sigma^2 I$, and Θ_i is estimated by minimizing the fitting error:

$$\epsilon_i = \|Y_i - \Phi_i \Theta_i\|$$

Then, two parts ω_1 and ω_2 obtained from two images of the same object with different, but similar, sensor and illumination configurations are considered appearances of the same part ω if

$$\epsilon_{1,2} = \frac{1}{n_1} \|Y_1 - \Phi_1 \Theta_2\| + \frac{1}{n_2} \|Y_2 - \Phi_2 \Theta_1\| \leq T_\epsilon$$

and

$$\Delta_{1,2} = \|\mu_1 - \mu_2\| \leq T_\Delta$$

where μ_1 and μ_2 are the centroids of the parts and T_ϵ and T_Δ are given thresholds. Note that these thresholds can be set according to the estimated noise covariance matrix $\sigma^2 I$ and the known difference in sensor locations. Furthermore, this criterion can handle both over- and under-segmentation problems by assigning more than one part in one frame to a part in the other frame.

Figures 2(a) and (c) show two images of the object “Lamp” when the camera is located at the positions corresponding to 20° and 30° . Figures 2(b) and (d) show the obtained parts, sorted in decreasing order of size. Finally, Table 1 gives the resulting correspondences (entries of 1) when the above criterion is used. Note that region 2 in the frame corresponding to 20° is correctly assigned to regions 2 and 6 in the frame corresponding to 30° .

2.3 Collection of Appearances

The effects of the sensor and illumination configurations on the appearance of a part are learned by collecting appearances of the same part in sequences of images under all possible configurations. Appearances of a part can be easily tracked through frames by using the matching criterion presented in the previous section. However, a tracking algorithm must also take into consideration that due to self-occlusion, and under- and over-segmentation problems, a part may disappear, split into several parts or merge with

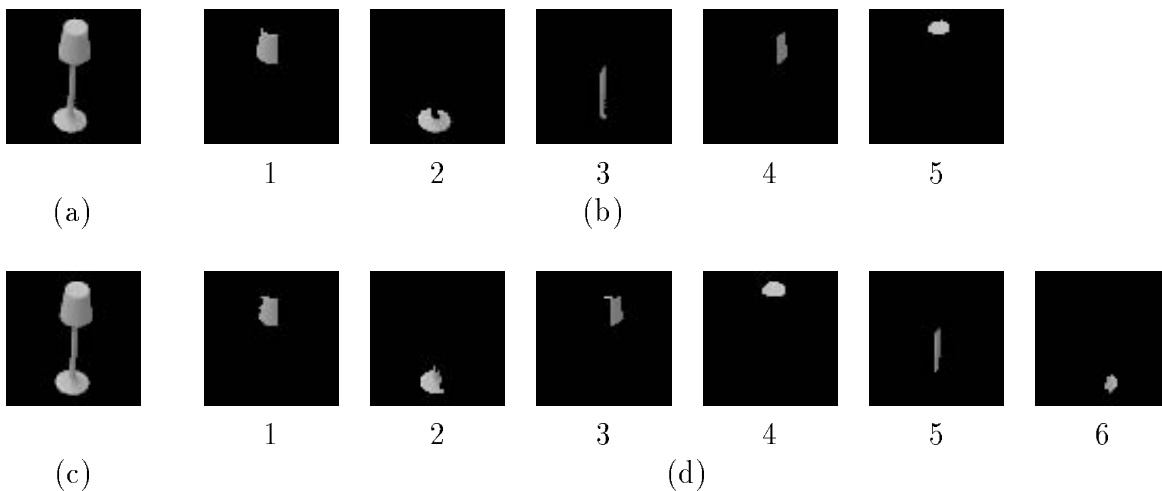


Figure 2: Appearances of parts. (a) Image of object “Lamp” with camera at 20° . (b) Parts segmented from (a). (c) Image of the same object with camera at 30° . (d) Parts segmented from (c).

Table 1: Part correspondences for object “Lamp” for views at 20° and 30° .

| Lamp Parts | | 30° | | | | | |
|---------------|---|------------|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| 20° | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| | 2 | 0 | 1 | 0 | 0 | 0 | 1 |
| | 3 | 0 | 0 | 0 | 0 | 1 | 0 |
| | 4 | 0 | 0 | 1 | 0 | 0 | 0 |
| | 5 | 0 | 0 | 0 | 1 | 0 | 0 |

others. Figure 3 represents a sequence of appearances of a part through ten different frames, f_0, f_1, \dots, f_9 . The numbers between the arrows in the figure correspond to the part size numbers in the different frames (the larger the number, the smaller the part), and the arrows link the appearances from one frame to the next. In this example, the part being tracked splits into two parts in frame f_3 , merges back into a single part in frame f_5 , only to split again in frame f_6 and to merge back in frame f_8 . Thus, it is fair to ask whether this part should be considered as one or two parts. We have chosen the majority rule criterion: if the number of frames where the tracked part is split is larger than half of the number of frames, it is decided that these are the appearances of two parts and that under-segmentation has occurred in the remaining frames; on the other hand, if the number of frames where the part is split is less than 50% of the frames, as in this example, it is decided that it is indeed a single part with over-segmentation occurring at the split frames. Note that whenever it is decided that there is a case of under-segmentation it is assumed that parts are being merged, and hence are sharing

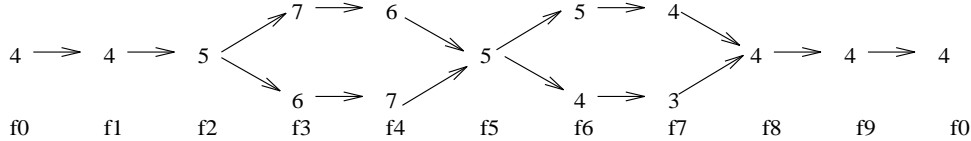


Figure 3: Example of splits and merges of appearances of a part through ten frames f_0, f_1, \dots, f_9 . The part being tracked splits into two in frame f_3 , merges back into a single part in frame f_5 , only to split again in frame f_6 and merge back in frame f_8 .

appearances in some of the frames.

Figures 4 and 5 illustrate the appearances of parts of two objects, “HoleCube” and “Lamp”. Figures 4(a) and 5(a) show images of these objects every 30° and Figures 4(b) and 5(b) show their respective MDL segmentations. Figures 4(c) and 5(c) show the appearances of five parts of each object. Note that due to self-occlusion, four of the parts of “HoleCube” disappear in some frames, and that due to segmentation problems the second and third parts of “Lamp” share appearances.

2.4 Appearance-Based Parts

The groups of appearances can be compactly stored and efficiently retrieved by constructing parametrized manifolds interpolating the projections of the individual appearances into eigenspaces obtained by applying the Karhunen-Loeve (K-L) compression method [31] to a scale- and brightness-normalized set of appearances.

Consider a collection of training appearances, Y_1, Y_2, \dots, Y_n , that have been scale- and brightness-normalized — i.e., their bounding boxes have been scaled to be $N = N_1 \times N_2$ pixels and their gray values have been scaled such that the $N \times 1$ vectors Y_i have unit length, $\|Y_i\| = 1$, $i = 1, \dots, n$. Let Q be the $N \times n$ matrix

$$Q = [Y_1 - \bar{Y} | Y_2 - \bar{Y} | \dots | Y_n - \bar{Y}]$$

where \bar{Y} is the mean value of the training appearances, and let S be the $N \times N$ covariance matrix

$$S = QQ^T$$

Then, using the K-L reduction method, an appearance Y can be expressed as a linear combination of $M \ll N$ eigenvectors of the covariance matrix S of the training appearances:

$$Y \sim \hat{Y} = \bar{Y} + EC$$

where E is an $N \times M$ matrix of the eigenvectors of the covariance matrix S with the largest M eigenvalues, and $C = E^T(Y - \bar{Y})$ is an $M \times 1$ vector of coefficients.

Since only a limited number of actual appearances are used during training, intermediate appearances are interpolated between them to obtain a more dense representation. These interpolated appearances, together with the projections of the training ones, form manifolds in the M -dimensional eigenspace spanned by E . These manifolds are like the ones proposed in [29], which have been shown to be successful when used to recognize

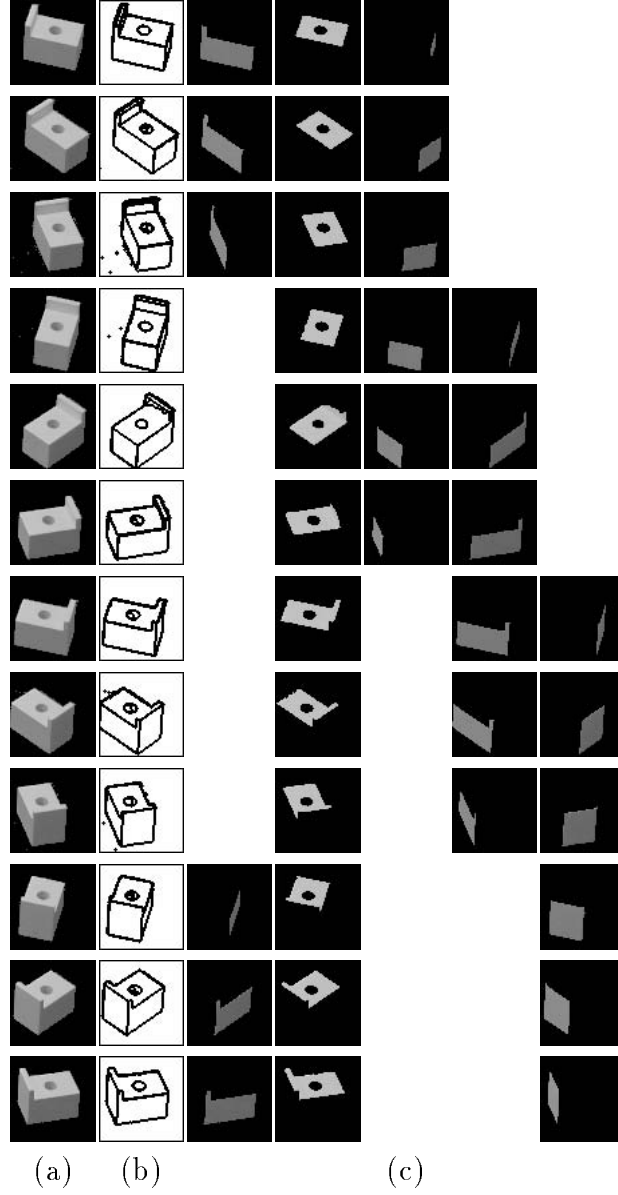


Figure 4: Collection of appearances of parts for “HoleCube”. Note that four of the parts disappear in some of the frames. (a) Images of “HoleCube” every 30° . (b) MDL segmentations of the images in (a). (c) Appearances of five parts.

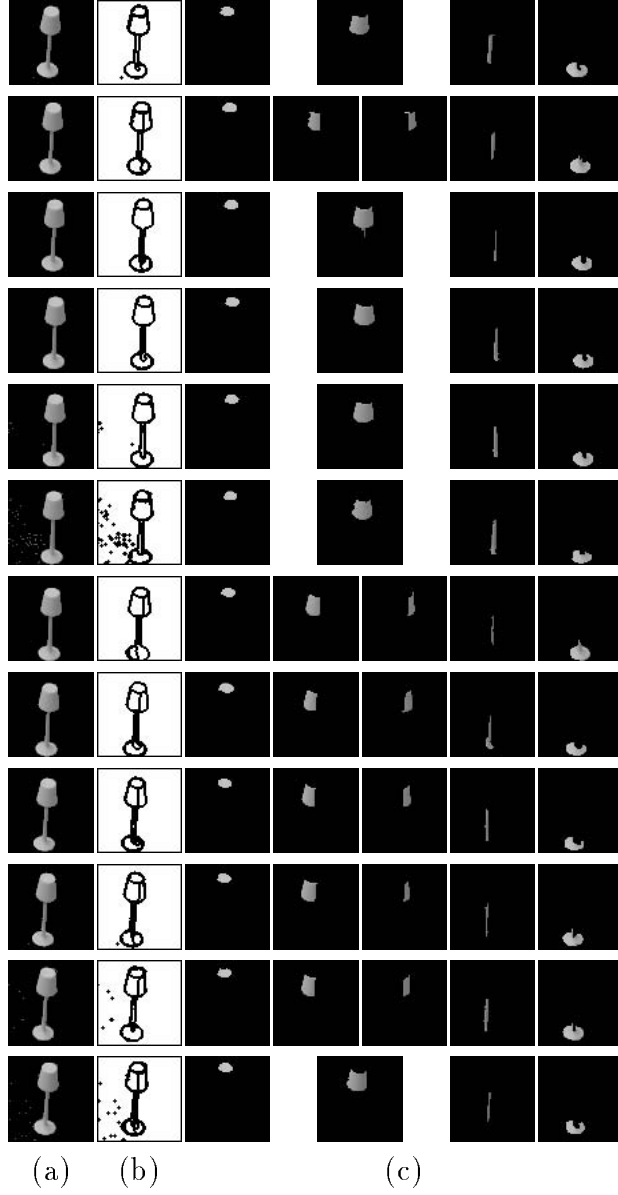
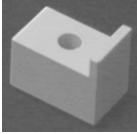




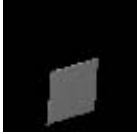


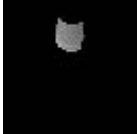










Figure 5: Collection of appearances of parts for “Lamp”. Note that the second and third part share appearances in some frames. (a) Images of “Lamp” every 30° . (b) MDL segmentations of the images in (a). (c) Appearances of five parts.

Table 2: ABP Database Sample. The ABPs of each object are represented by one of their appearances.

| Object | ABPs Representatives | | | | | |
|---|---|---|---|--|---|--|
|  |  |  |  |  |  | |
|  |  |  |  |  | | |
|  |  |  |  |  |  | |

and locate isolated objects. However, until now they have been used to represent appearances of complete objects and therefore have failed in the presence of occlusion. In this paper, we propose to use this type of representation with parts, taking advantage of their good localization properties while addressing the occlusion problem. Formally, we define appearance-based parts:

An **appearance-based part** (ABP) is a parametrized manifold in a space spanned by a given set of scale- and brightness-normalized appearances of parts, representing a collection of appearances of a part, obtained by varying the viewing conditions within a given space.

ABPs can be easily constructed with the software package SLAM [30] developed at Columbia University; it only requires having 1) a set of appearances of parts spanning an eigenspace; and 2) a collection of appearances of parts to obtain the corresponding manifold. The set used to span the eigenspace can be chosen in many ways. It can be, for example, the set of all the collections of appearances of parts for a single or several objects. Table 2 shows representative appearances for the ABPs of three objects, where each of these ABPs has an associated manifold in the ABP space. Figure 6 (b) shows the first five eigenvectors of the space spanned by compressing the appearances of the five parts collected from the object “HoleCube” shown in Figure 6 (a), sorted in decreasing order of the magnitude of the corresponding eigenvalue. Figure 6 (c) shows an appearance of one of the parts of this object, and Figure 6 (d) shows the manifold obtained by collecting appearances of this part, displayed three-dimensionally, for visualization purposes.

2.5 Appearance-Based Relationships

Although it may be possible to identify some objects by recognizing some of their distinctive parts, recognizing most objects requires the use of relations among the parts. This

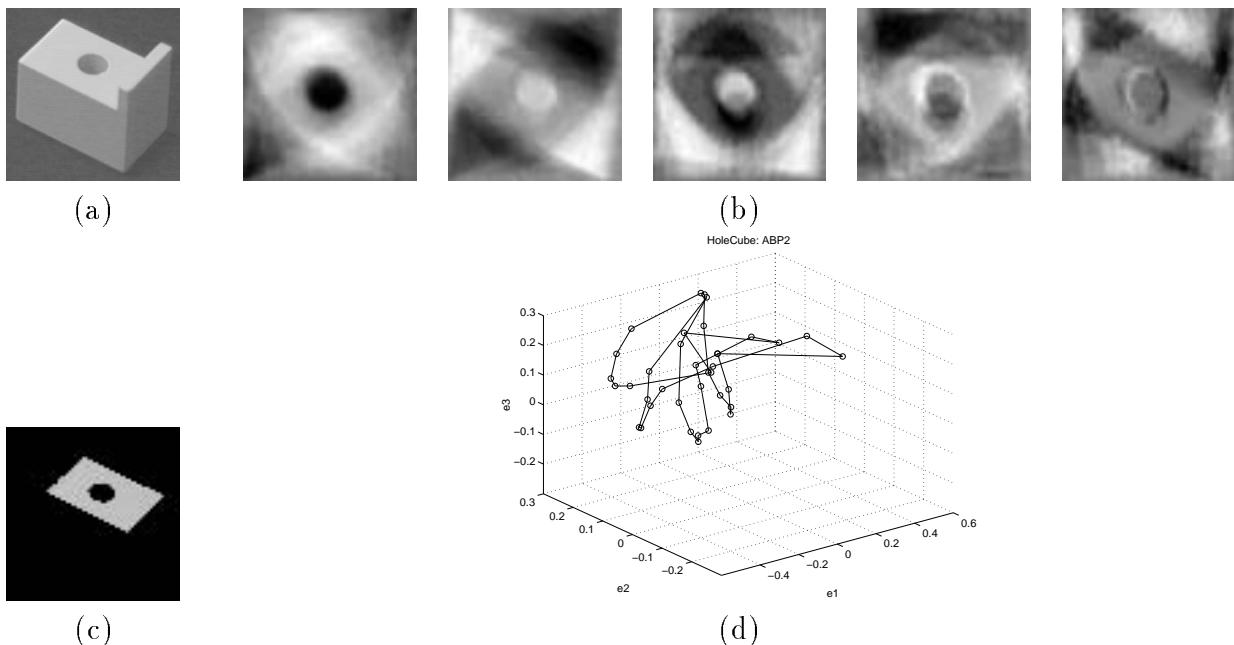


Figure 6: ABP Representation. (a) Object “HoleCube.” (b) First five eigenvectors of the space spanned by the ABPs of “HoleCube,” sorted in decreasing order of magnitude of the eigenvectors. (c) One appearance of a part of (a). (d) Manifold corresponding to the collection of appearances of the parts shown in (c), displayed three-dimensionally for visualization purposes.

structure can be expressed using attributed relational descriptions [39], reducing the recognition task to finding an isomorphism or partial isomorphism between the model and image graphs [41, 15, 16, 40].

The graph representation is well suited to describe relational structure when the relation is *static* or *quasi-static* — i.e. when attributes and relations hold for a sufficiently large set of viewing conditions² [11]. However, relations between ABPs are not static, since their attributes as well as their structure change with the viewing conditions. For example, two ABPs may be adjacent to each other only in a subset of all possible viewing conditions. Furthermore, their size and their length of common boundary may change for different viewing conditions within this subset. Thus, attributed relational graphs using ABPs must change accordingly, making a traditional graph representation cumbersome. A solution to this problem is to decompose the graph into sub-graphs of relations that hold for a subset of views and to capture the changes of their attributes and the attributes of the parts involved (due to variations in viewing conditions) in manifolds similar to the ones used to describe parts, as explained next.

The reason for using sub-graphs is two-fold: First, sub-graphs restrict the relations to subsets of parts and make the range of viewing conditions for which relations hold larger. Second, sub-graphs are a natural choice when performing recognition in the presence of

²An example of a static relation is the adjacency between the legs and top of a table.

occlusion [5, 11]. Consider two extreme cases: in one case, all of the object except for one part is occluded; in the other, all of the object is visible. Between these extremes, a number of cases are possible depending on how many and which of the parts of the object are visible. While storing all the possible sub-graphs is prohibitive due to the combinatorics involved, it is possible to store subsets of them, such as pairs and triples of adjacent parts³. These cases correspond to increasingly larger regions of the object being visible, and can be considered both as higher-level features and as spatial relationships between lower-level features. In principle, these more complex features are more sensitive to occlusion (since they require more than one part being visible). However, they have higher discriminatory power and provide intermediate levels of representation between global and atomic approaches.

Once the subsets of viewing conditions for which the relations hold are identified, it is possible to apply the Karhunen-Loeve compression method to subimages containing only the parts involved in the relations. These subimages can then be compactly stored and efficiently retrieved by constructing parametrized manifolds interpolating their projections in a lower-dimensional eigenspace. These manifolds capture the appearance of the relations for the different views within the sets for which they hold, and thus are called *Appearance-Based Relations* (ABRs). Table 3 shows representative appearances for the ABRs of three objects when the adjacency relation between pairs of regions is used.

3 Object Recognition

The ABPs and ABRs described above can be used in a relational matching framework as the basis for an object recognition system. Let \mathcal{ABP} and \mathcal{ABR} represent the sets of ABPs and ABRs, respectively, for all the objects in a given database. Then the object database can be represented by a set of relational descriptions

$$\mathcal{DB} = \{\mathcal{D}_1, \dots, \mathcal{D}_N\}$$

where

$$D_m = \{R_1, R_2\}$$

is the relational description for object m ,

$$\mathcal{R}_1^m = \{P_1, \dots, P_{m_1}\} \subseteq \mathcal{ABP}$$

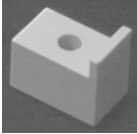









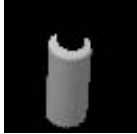












is the set of all the ABP manifolds, P_i , $i = 1, \dots, m_1$, belonging to model m , and

$$\mathcal{R}_2^m = \{R_1, \dots, R_{m_2}\} \subseteq \mathcal{ABR}$$

is the set of all the ABR manifolds, R_i , $i = 1, \dots, m_2$, belonging to model m . The set \mathcal{R}_1^m is a unary relation of parts, while the set \mathcal{R}_2^m is a binary relation between parts — i.e. $R_i = (P_{i_1}, P_{i_2})$, $i = 1, \dots, m_2$, $1 \leq i_1, i_2 \leq m_1$. For example, the relational description of the object “HoleCube” is formed by a relation R_1 comprised of all the ABPs shown in

³This is analogous to the use of junctions of curves in approaches using curve segments as the main representation feature.

Table 3: ABR Database Sample. The ABRs of each object are represented by one of their appearances.

| Object | ABRs Representatives | | | | | |
|---|--|--|--|---|---|--|
|  |  |  |  |  |  | |
| |  |  |  | | | |
|  |  |  |  | | | |
|  |  |  |  |  |  | |
| |  |  |  |  | | |

the first row of Table 2 and a relation R_2 comprised of all the ABRs shown in the first row of Table 3.

An MDL segmentation of an image can be described using a similar relational representation

$$D_i = \{S_1, S_2\}$$

where

$$\mathcal{S}_1 = \{p_1, \dots, p_{n_1}\}$$

is a unary relation formed by a set of projections of parts or image regions into the ABP eigenspace, and where

$$\mathcal{S}_2 = \{r_1, \dots, r_{n_2}\}$$

is a binary relation formed by a set of projections of pairs of adjacent parts into the ABR eigenspace.

The main difference between these representations is that the description of an object is made in terms of ABPs and ABRs — i.e., collections of appearances — while the description of an image segmentation is made of a particular instance of these appearances.

An image is an *observation* of a subset of the models. Then, the *recognition* problem is to find two unknown correspondence mappings

$$h_{ABP} : \mathcal{S}_1 \rightarrow \mathcal{ABP}$$

$$h_{ABR} : \mathcal{S}_2 \rightarrow \mathcal{ABR}$$

associating ABPs and ABRs with image regions and pairs of regions, respectively, and the *localization* problem is to find two unknown correspondence mappings

$$\begin{aligned} l_{ABP} &: \mathcal{S}_1 \rightarrow \mathcal{R}_1^m \\ l_{ABR} &: \mathcal{S}_2 \rightarrow \mathcal{R}_2^m \end{aligned}$$

associating appearances of ABPs and ABRs with image regions and pairs of image regions, respectively.

The mappings h_{ABP} and h_{ABR} represent a set of ABP and ABR *identity* hypotheses while the mappings l_{ABP} and l_{ABR} represent a set of ABP and ABR *pose* hypotheses. These hypotheses constrain each other and can be generated by projecting each segmented (or pair of adjacent) region(s) into the ABP (ABR) eigenspace, and finding the closest points on the closest manifolds to this projection. While the closest manifolds provide hypotheses for the identity mapping, the closest point on each manifold provides a hypothesis for the localization problem.

Let r be the projection of a pair of adjacent image regions with projections p_1 and p_2 and let $h_{ABR}(r) = R \in \mathcal{R}_2^m$ and $l_{ABR}(r) = a_R \in R$ be its ABR identity and pose hypotheses, respectively. If the ABP hypotheses for p_1 and p_2 , $h_{ABP}(p_1) = P_1$ and $h_{ABP}(p_2) = P_2$, are such that $P_1, P_2 \in \mathcal{R}_1^m$ and $R = (P_1, P_2)$ we say that the ABR hypothesis for r is *compatible* or verifies the ABP hypotheses for p_1 and p_2 . Furthermore, if the ABP identity hypotheses are compatible with the ABR hypothesis *and* the ABP pose hypotheses for p_1 and p_2 , $l_{ABP}(p_1) = a_{P_1} \in P_1$ and $l_{ABP}(p_2) = a_{P_2} \in P_2$, are such that a_{P_1} and a_{P_2} correspond to the same pose, we say that the ABR pose hypothesis for r is compatible with or verifies the ABP pose hypotheses for p_1 and p_2 .

Finally, let $d(p, q)$ represent a distance metric between two points p and q in a given eigenspace and let the distance between a point p and a manifold M be defined as the distance between the point p and the closest point to p on the manifold, $d(p, M) = \min_{q \in M} d(p, q)$. Then, the distances between the projections of the image regions and pairs of regions and the corresponding manifolds and appearances $d(p, h_{ABP}(p))$, $d(r, h_{ABR}(r))$, $d(r, l_{ABR}(r))$, and $d(p, l_{ABP}(p))$ are quantitative measures of the goodness of these hypotheses, where the smaller the distance, the better the match.

4 Experiments and Results

In this section we describe a set of experiments to illustrate the potential of the proposed representation for successful object recognition in the presence of occlusion and clutter. For these experiments, a very simple-minded strategy was used to generate hypotheses: those ABP hypotheses with distance $d = d(p, h_{ABP}(p)) \leq T_1$, where T_1 is a “small” threshold, were taken as successful hypotheses. Other ABP hypotheses with somewhat larger distances $T_1 \leq d \leq T_2$, where T_2 is a second threshold, were verified or discarded by using ABR hypotheses. Finally, an ABR hypothesis for a pair of image parts $r = (p_1, p_2)$ was said to verify the ABP hypotheses for the component parts p_1 and p_2 if the distance $d(r, h_{ABR}(r)) \leq T_3$, where T_3 is a third threshold.

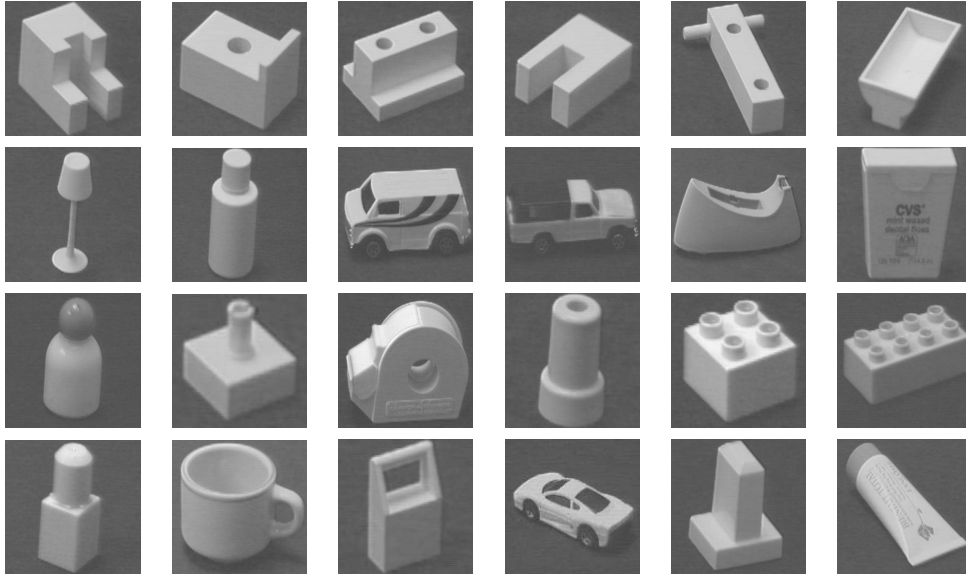


Figure 7: Object Database.

Figure 7 shows images of the objects in our current database. The ABP database corresponding to these objects has a total of 110 ABPs and the ABR database has a total of 130 ABRs.

4.1 Qualitative Results

Examples of cluttered scenes with busy backgrounds are shown in Figure 8. The first column shows the original images, the second column shows their MDL segmentations, and the following columns show the appearances of the ABPs and ABRs that were hypothesized and verified by the above strategy, shown at the hypothesized pose. Whenever an ABP is shown, the object was identified based only on a “distinctive” ABP ($d(p, h_{ABP}(p)) \leq T_1$). Whenever an ABR is shown, the object was identified by verifying two ABP hypotheses using an ABR ($T_1 \leq d(p_1, h_{ABP}(p_1)), d(p_2, h_{ABP}(p_2)) \leq T_2$; $d((p_1, p_2), h_{ABR}((p_1, p_2))) \leq T_3$). It is seen that, in spite of the occlusion between the objects and segmentation problems such as the merging of some of the object parts with the background, all the objects and their poses are correctly identified.

4.2 Quantitative Results

While the examples shown above provide anecdotal evidence that our representation is capable of handling occlusion and segmentation problems, they do not provide quantitative data to characterize its performance. The most common tool used to present data characterizing the performance of a detection algorithm is a plot of its probability of misdetection versus its probability of false alarm, as some tuning parameter (threshold) is varied [22]. This plot is commonly known as the “receiver operating curve” or ROC

for short. Next, we describe two sets of experiments that were designed to obtain data to characterize the performance of ABP and ABR detection using ROCs. In the first set of experiments, the effect of occlusion on the recognition of individual parts was studied. In the second set of experiments, the performance of the recognition strategy was quantified for a set of real scenes, as the three thresholds T_1 , T_2 , and T_3 were varied.

4.2.1 Performance Characterization of ABP and ABR Recognition

Experiments to characterize the effect of occlusion on ABP and ABR recognition were conducted by randomly selecting an appearance of an ABP or ABR from the database, and randomly occluding a percentage of its bounding box area with the appearance of another ABP, also randomly selected. Figure 9 illustrates the synthetic occlusion generation applied to an appearance of an ABP.

The projection of the occluded region(s) was assigned the closest manifold in the database. If the distance between this projection and the manifold was larger than a threshold T , it was said that the experiment resulted in a *misdetction*. If this distance was less than the threshold T , but the assigned manifold was not the one corresponding to the true (known) identity of the region(s) being used, it was said that this was a *false alarm*. Otherwise, it was said that the projection was correctly identified. Figures 10 (a) and (b) show plots of false alarms versus misdetections for ABPs and ABRs, respectively. The experiments were done for 10,000 occluded pairs, with occlusions ranging between 10% and 50%, as the threshold T varied from 0.015 to 0.15. It is seen that ABPs and ABRs are affected in a similar way when the occlusion is small (10%). On the other hand, ABPs are more sensitive than ABRs when the percentage of occlusion increases, and this sensitivity increases with the level of occlusion. This can be explained by the fact that the shapes of ABRs tend to be more irregular and distinctive than the shapes of ABPs, and thus are easier to identify when the amount of occlusion increases. It should also be noted that the performance could be significantly improved by obtaining the region projections using a robust technique such as the one proposed in [26] instead of using least squares as we did.

4.2.2 Performance Characterization of the Recognition Strategy

Figures 11 (a) and (b) show images of two scenes with three objects from the database, set up on top of a rotating table. In order to study the performance of the recognition strategy, twelve images of each scene, from different points of view, were taken by rotating the table in increments of 30 degrees. Let C be the number of extracted image parts that belong to the clutter but are incorrectly identified as object parts, and D and M be the number of parts that belong to one of the objects in the database and that are correctly identified or missed by the recognition strategy, respectively. The system performance can be characterized in terms of recall and precision⁴. The system *recall* is measured by the ratio between the number of ABPs correctly identified and the number of ABPs

⁴Precision and recall are standard evaluation metrics used in the information retrieval community [37].

present in the image:

$$\text{Recall} = \frac{D}{M + D}$$

and the *misdetecion rate* is defined as the ratio

$$MD = \frac{M}{M + D} = 1 - \text{Recall}.$$

The system *precision*, on the other hand, is measured by the ratio between the number of correctly identified object parts and the total number of identified parts:

$$\text{Precision} = \frac{D}{C + D}$$

and the *false alarm rate* is defined as the ratio

$$FA = \frac{C}{C + D} = 1 - \text{Precision}.$$

Figure 11 (c) shows plots of false alarms vs. misdetections of ABPs as the thresholds T_1, T_2 and T_3 are varied. The plot for $T_2 = T_3 = 0$ shows the performance of the system as T_1 varies between 0 and 0.08, when only ABPs are used. In this case, the best performance is attained for $T_1 = 0.03$ with a false alarm rate of 0.098 and a misdetection rate of 0.111. Setting T_2 to the value corresponding to 0 ABP misdetections, 0.08, and using ABRs to verify ABP hypotheses with a threshold $T_3 = 0.05$ improves the performance to a false alarm rate of 0.103 and a misdetection rate of 0.093, when $T_1 = 0.03$. Finally, setting both T_2 and T_3 to 0.08 further improves the performance to a false alarm rate of 0.082 and a misdetection rate of 0.085, with $T_1 = 0.02$. Thus, it is seen that a simplistic recognition strategy using ABPs and ABRs, like the one suggested here, results in relatively low false alarm and misdetection rates (less than 10%). This performance can be improved by using more sophisticated strategies, such as allowing multiple hypotheses to compete in a Bayesian framework [7], and using robust projection algorithms [26] to generate ABP and ABR hypotheses.

5 Conclusion

In this paper we introduced a new object representation using appearance-based parts (ABP) and relations (ABR). ABPs and ABRs are defined based on the MDL principle and are automatically learned from collections of images without requiring *ad hoc* parameters. They capture not only local shape but also intrinsic reflectance properties, pose in the scene and illumination conditions. Furthermore, ABPs and ABRs are compactly stored using an eigenspace representation parametrized by pose and illumination. Thus, the proposed representation can be used with generic objects and it is robust to occlusion and segmentation variations. The usefulness of the representation was illustrated by qualitative as well as quantitative experiments. Qualitative examples showed that the representation can be used to identify and localize objects in scenes with significant levels of clutter and partial object occlusion, in spite of segmentation errors. Quantitative

results for the sensitivity of the ABP and ABRs to occlusion and their usefulness for object recognition were provided using receiver operating curves (ROCs). The sensitivity of the ABPs and ABRs to occlusion was characterized as the level of occlusion was increased. It was seen that, in general, ABRs are less sensitive to occlusion than ABPs. The performance of a simplistic object recognition strategy based on ABPs and ABRs and three thresholds was also characterized as the values of these thresholds were varied. It was shown that the misdetection and false alarm rates are lower when ABPs and ABRs are used instead of only ABPs, thus illustrating the power of the representation, and in particular the use of the appearance-based relations. Finally, since the experiments described here used a very simplistic recognition strategy, it is expected that better performance could be attained by using a more sophisticated recognition strategy where multiple hypotheses are allowed to compete in a Bayesian framework [7].

6 Acknowledgments

The authors would like to thank Dr. Nayar and Mr. Nene of Columbia University for providing the SLAM software library and their help in using it; Dr. Kao for some of the test objects; and Dr. Rosenfeld of the University of Maryland for comments. C. Y. Huang and O. I. Camps were supported in part by NSF grants IRI-93-09100 and IRI-97-12598. T. Kanungo was supported in part by the Department of Defense and the Army Research Laboratory under Contract MDA 9049-6C-1250.

References

- [1] N. Abramson. *Information Theory and Coding*. McGraw-Hill, 1963.
- [2] F. Arman and J. K. Aggarwal. CAD-based object recognition in range images using pre-compiled strategy trees. In A. K. Jain and P. Flynn, editors, *Three-Dimensional Object Recognition Systems*, pages 115–134. Elsevier Science Publishers, 1993.
- [3] T. O. Binford. Body-centered representation and perception. In *Lecture Notes in Computer Science (994): Object Representation in Computer Vision*. Springer-Verlag, 1995.
- [4] H. Bischof and A. Leonardis. Robust recognition of scaled eigenimages through a hierarchical approach. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 664–670, Santa Barbara, CA, June 1998.
- [5] R. C. Bolles and R. A. Cain. Recognizing and locating partially visible objects: The local-feature-focus method. *Int. Journal of Robotics Research*, 1(3):57–82, 1982.
- [6] O. I. Camps. Towards a robust physics-based object recognition system. In *Lecture Notes in Computer Science (994): Object Representation in Computer Vision*. Springer-Verlag, 1995.
- [7] O. I. Camps, C. Y. Huang, and T. Kanungo. Hierarchical organization of appearance based parts and relations for object recognition. In *Proc. IEEE Conference on*

- Computer Vision and Pattern Recognition*, pages 685–691, Santa Barbara, CA, June 1998.
- [8] O. I. Camps, L. G. Shapiro, and R. M. Haralick. Image prediction for computer vision. In A. K. Jain and P. Flynn, editors, *Three-dimensional Object Recognition Systems*. Elsevier Science Publishers, 1993.
 - [9] O. I. Camps, L. G. Shapiro, and R. M. Haralick. A probabilistic matching algorithm for computer vision. *Annals of Mathematics and Artificial Intelligence*, 10(1-2):85–124, 1994.
 - [10] C. Chen and P. Mulgaonkar. Automatic vision programming. *CVGIP: Image Understanding*, 55(2):170–183, 1992.
 - [11] M. S. Costa and L. G. Shapiro. Scene analysis using appearance-based models and relational indexing. In *International Symposium on Computer Vision*, pages 103–108, Coral Gables, FL, November 1995.
 - [12] S. J. Dickinson. Part-based modeling and qualitative recognition. In A. K. Jain and P. Flynn, editors, *Three-Dimensional Object Recognition Systems*, pages 201–228. Elsevier Science Publishers, 1993.
 - [13] S. J. Dickinson, A. P. Pentland, and A. Rosenfeld. 3D shape recovery using distributed aspect matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(2):174–198, 1992.
 - [14] W. E. L. Grimson, T. L. Poggio, S. J. White, and N. Noble. Recognizing 3D objects using constrained search. In A. K. Jain and P. Flynn, editors, *Three-Dimensional Object Recognition Systems*, pages 259–284. Elsevier Science Publishers, 1993.
 - [15] R. M. Haralick and L. G. Shapiro. The consistent labeling problem I. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1(2):173–184, 1979.
 - [16] R. M. Haralick and L. G. Shapiro. The consistent labeling problem II. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2(3):193–203, 1980.
 - [17] M. Hebert, J. Ponce, T. Boult, and A. Gross. Report on the 1995 Workshop on 3D Object Representations in Computer Vision. In *Lecture Notes in Computer Science (994): Object Representation in Computer Vision*. Springer-Verlag, 1995.
 - [18] C. Y. Huang, O. I. Camps, and T. Kanungo. Object recognition using appearance-based parts and relations. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 877–883, San Juan, Puerto Rico, June 1997.
 - [19] D. Huttenlocher. Recognition by alignment. In A. K. Jain and P. Flynn, editors, *Three-Dimensional Object Recognition Systems*, pages 311–326. Elsevier Science Publishers, 1993.

- [20] A. K. Jain and P. J. Flynn, editors. *Three-Dimensional Object Recognition Systems*. Elsevier, 1993.
- [21] T. Kanungo, B. Dom, W. Niblack, and D. Steele. A fast algorithm for MDL-based multi-band image segmentation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 609–616, Seattle, WA, June 1994.
- [22] T. Kanungo, M. Y. Jaisimha, J. Palmer, and R. Haralick. A methodology for quantitative performance evaluation of detection algorithms. *IEEE Trans. on Image Processing*, 4(12):1667–1674, 1995.
- [23] D. Kriegman and J. Ponce. Representations for recognizing complex curved 3D objects. In *Lecture Notes in Computer Science (994): Object Representation in Computer Vision*. Springer-Verlag, 1995.
- [24] J. Krumm. Eigenfeatures for planar pose measurement of partially occluded objects. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 55–60, San Francisco, CA, June 1996.
- [25] Y. Leclerc. Region grouping using the minimum description length principle. In *Proc. DARPA Image Understanding Workshop*, pages 473–479, Pittsburgh, PA, Sept. 1990.
- [26] A. Leonardis and H. Bischof. Dealing with occlusions in the eigenspace approach. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 453–458, San Francisco, CA, June 1996.
- [27] D. Metaxas. A physics-based framework for segmentation, shape and motion estimation. In *Lecture Notes in Computer Science (994): Object Representation in Computer Vision*. Springer-Verlag, 1995.
- [28] J. Mundy, A. Liu, N. Pillow, A. Zisserman, S. Abdallah, S. Utcke, S. Nayar, and C. Rothwell. An experimental comparison of appearance and geometric model based recognition. In *Lecture Notes in Computer Science (1144): Object Representation in Computer Vision II*. Springer-Verlag, 1996.
- [29] H. Murase and S. K. Nayar. Visual learning and recognition of 3D objects from appearance. *International Journal of Computer Vision*, 14(1):5–24, 1995.
- [30] S. Nene, S. K. Nayar, and H. Murase. *SLAM: Software Library for Appearance Matching*. Technical Report CUCS-019-94, Department of Computer Science, Columbia University, 1994.
- [31] E. Oja. *Subspace methods of Pattern Recognition*. Research Studies Press, Hertfordshire, UK, 1983.
- [32] A. Pentland and S. Sclaroff. Closed-form solutions for physically based shape modeling and recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(7):715–729, 1991.

- [33] J. Ponce, A. Zisserman, and M. Hebert. Report on the 1996 International Workshop on Object Representation in Computer Vision. In *Lecture Notes in Computer Science (1144): Object Representation in Computer Vision II*. Springer-Verlag, 1996.
- [34] A. R. Pope and D. G. Lowe. Learning appearance models for object recognition. In *Lecture Notes in Computer Science (1144): Object Representation in Computer Vision II*. Springer-Verlag, 1996.
- [35] J. Rissanen. A universal prior for integers and estimation by minimum description length. *The Annals of Statistics*, 11(2):211–222, 1983.
- [36] C. A. Rothwell. *Object Recognition through Invariant Indexing*. Oxford University Press, 1995.
- [37] G. Salton and M. E. Lesk. Computer evaluation of text indexing and text processing. *Journal of the Association for Computing Machinery*, 15(1):8–36, 1968.
- [38] S. Sclaroff and A. P. Pentland. Modal matching for correspondence and recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(6):545–561, 1995.
- [39] L. Shapiro, J. D. Moriarty, R. M. Haralick, and P. G. Mulgaonkar. Matching three-dimensional objects using a relational paradigm. *Pattern Recognition*, 17(4):385–405, 1984.
- [40] L. G. Shapiro and R. M. Haralick. Structural descriptions and inexact matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 3(5):504–519, 1981.
- [41] J. R. Ullmann. An algorithm for subgraph isomorphism. *Journal of the Association for Computing Machinery*, 23(1):31–42, 1976.
- [42] M. Zerroug and G. Medioni. The challenge of generic object representation. In *Lecture Notes in Computer Science (994): Object Representation in Computer Vision*. Springer-Verlag, 1995.
- [43] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(9):884–900, 1996.
- [44] A. Zisserman, D. Forsyth, J. Mundy, C. Rothwell, J. Liu, and N. Pillow. 3D object recognition using invariance. *AI Journal*, 78(1-2):239–288, 1995.

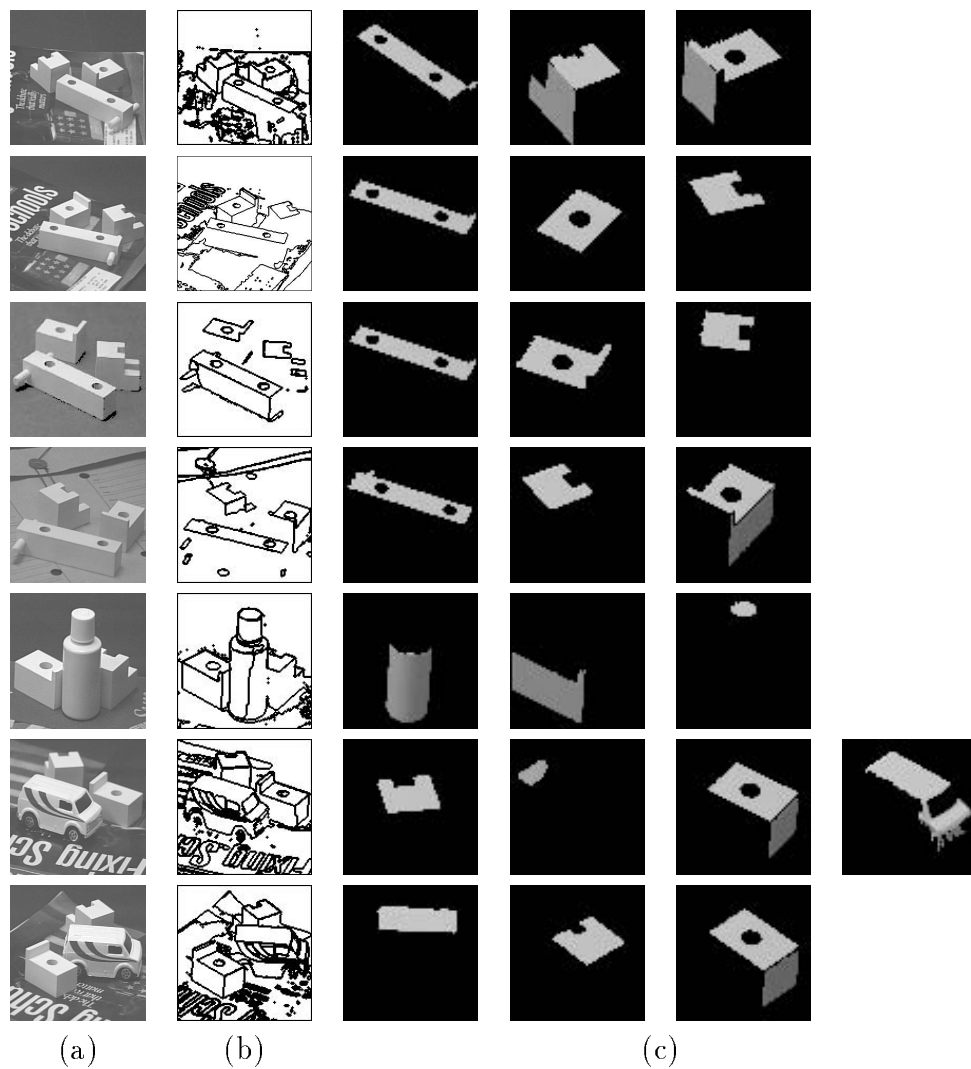


Figure 8: Results for cluttered scenes. (a) Cluttered scenes. (b) MDL segmentations. (c) ABP and ABR hypotheses.

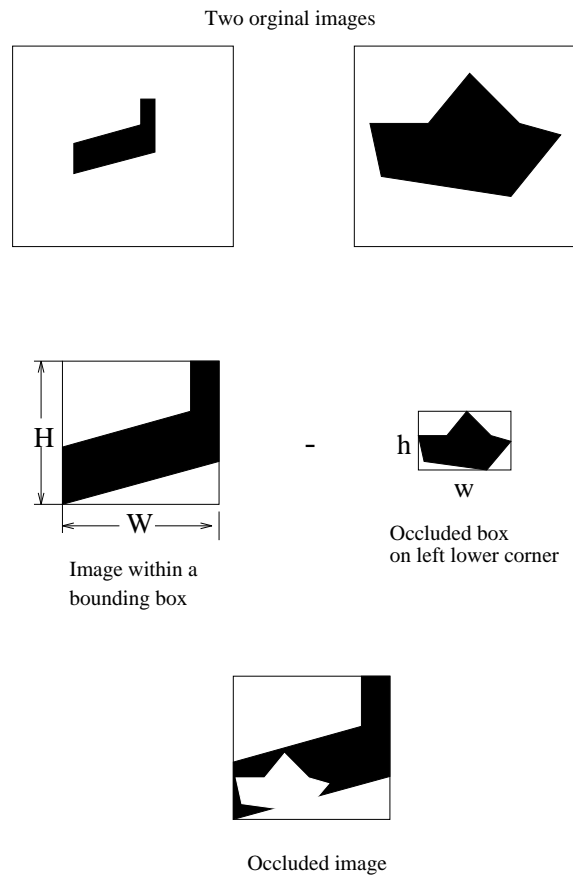
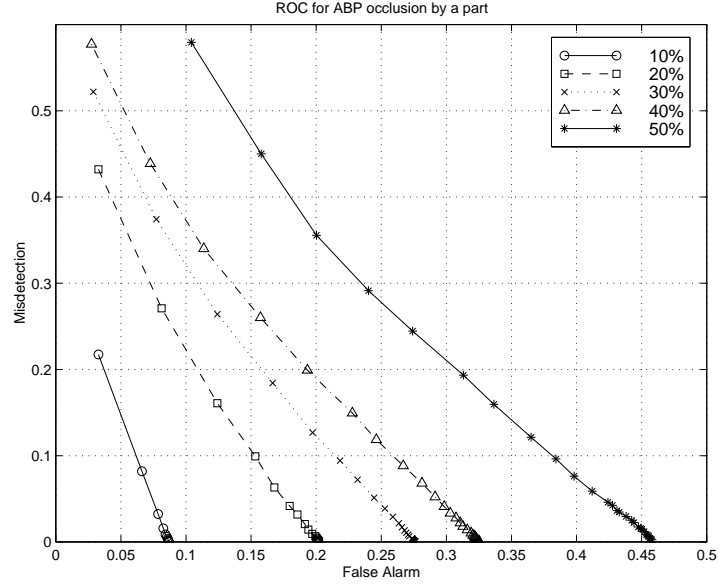
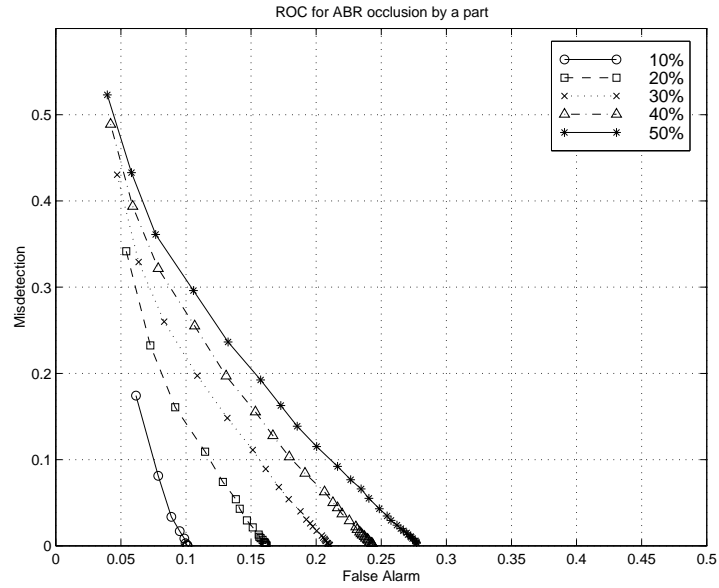


Figure 9: Image occlusion process using another image frame in the database.

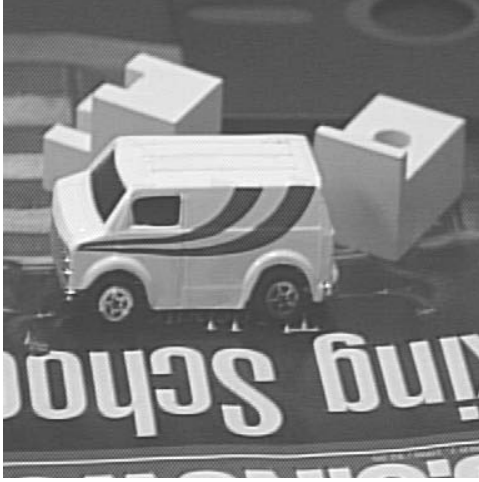


(a)



(b)

Figure 10: (a) False alarm and misdetection of ABPs when occluded by another part, for different levels of occlusion. (b) False alarm and misdetection of ABRs when occluded by a part, for different levels of occlusion.



(a)



(b)

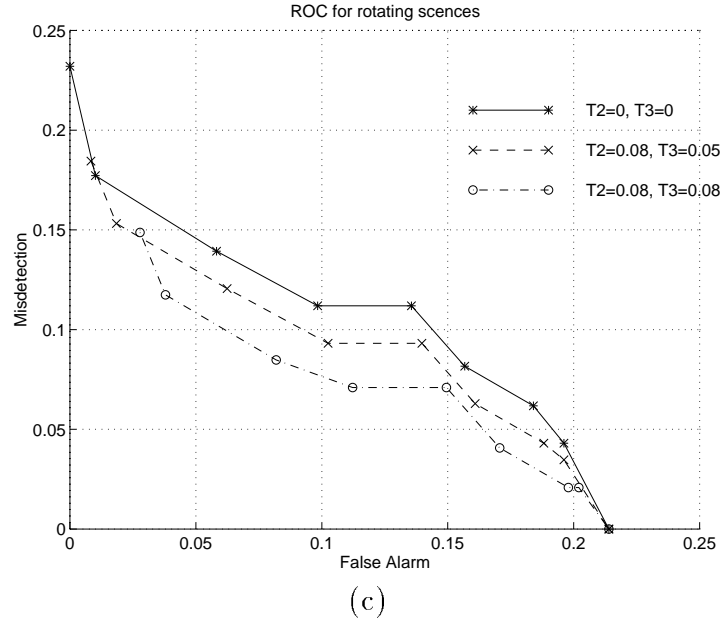


Figure 11: False alarm versus misdetections for rotating scenes shown in (a) and (b). (c) False alarm and misdetction when T_1, T_2 and T_3 are varied. It is seen that using ABRs to verify ABP hypotheses ($T_2 = 0.08, T_3 = 0.05$ and $T_2 = 0.08, T_3 = 0.08$) results in better performance than using ABPs alone ($T_2 = 0, T_3 = 0$).